ESAIM23 – 1st European Symposium on Artificial Intelligence in Manufacturing

Vision-Based Ladle Monitoring System for Steel Factories

Mohamed Selim[1,*], Pablo López de Uralde[2], Jon Mata[2], Eider Gorostegui-Colinas[2], Beatriz Chicote[2], Alain Pagani[1], and Didier Stricker[1]

[1] German Research Center for Artificial Intelligence, DFKI GmbH, Germany
[2] Lortek, Spain

[*] Corresponding author. E-mail: mohamed.selim@dfki.de

Abstract. This paper presents a vision-based ladle monitoring system for steel factories, consisting of two modules: one for ladle surface temperature analysis using thermal cameras and another for deep learning-based detection and recognition of ladle identification numbers. The first module monitors ladle thermal behavior by capturing high-resolution thermal images and employing advanced image analysis techniques. This enhances safety and efficiency in steel production. The second module focuses on digit recognition on the ladle surface, providing crucial identification and tracking information. A robust deep learning model trained on a large dataset of thermal camera images achieves high accuracy in ladle identification. The proposed system integrates thermal cameras and advanced image analysis techniques, offering real-time monitoring, early anomaly detection, and accurate ladle identification. Experimental evaluations demonstrate its effectiveness, indicating its potential for practical implementation in steel factory environments.

Keywords: Ladle monitoring · Thermal cameras · Deep Learning · Steel manufacturing.

1 Introduction

A ladle is a steel vessel used for transporting liquid steel in a factory. Efficient coordination of ladle movements is crucial to minimize heat losses and ensure optimal supply. The lifetime of a ladle depends on refractory lining wear, which can have significant economic and safety impacts if not addressed promptly. Premature retirement of ladles increases refractory costs and overall production expenses. Currently, ladle lifetime decisions rely on visual observations by experienced personnel. However, existing commercial laser measuring systems lack precision in assessing the remaining refractory layer, mainly focusing on major cracks. Quality monitoring using cameras in the industry has revolutionized manufacturing processes, enabling real-time visual inspections that enhance product consistency and minimize defects [6], [5]. These advanced camera systems not only detect imperfections but also provide valuable data for process optimization, ensuring higher standards of quality control. Thermal cameras offer the ability to monitor ladle surface temperature, detect refractory wear-related issues, and track temperature changes. This allows for the identification of hotspots near breakout risks and the prediction of future ladle behavior by analyzing temperature evolution along different profiles. As ladle parts experience varying thermal history and wear, thermal monitoring provides a comprehensive evaluation of both current and future ladle states.

Considering the cycle of a ladle from the Electric Arc Furnace (EAF) to the casting station, the ladle is tracked and identified by the factory system. However, tracking and identification are not accurately performed at the cleaning and preheating stations. As a result, the time spent by the ladles at these stations is not precisely calculated. This time information is valuable for planning ladle usage in heats. If a ladle spends a short time after cleaning at the burners, it requires less preheating before it can be used in the next heat. Therefore, ladle identification at these stations is important. Each ladle in the steel factory has a specific number written on its body using steel pieces. The ladle number is welded on the surface, typically in multiple locations. Computer vision techniques can be employed to locate and identify the ladle number accurately. However, the harsh environment at the steel factory, including dust, can obstruct vision. Therefore, careful consideration and study are required when choosing cameras for this task. Color and thermal images have been explored for ladle tracking, each with its own advantages and disadvantages compared to the other.

The innovation of the proposed method lies in its comprehensive approach to ladle monitoring within steel factories. By combining thermal cameras and advanced image analysis techniques, this system not only tracks ladle surface temperatures with precision but also achieves digit recognition for ladle identification numbers through deep learning. This integrated approach enhances safety and efficiency in steel production, offering real-time monitoring, early anomaly detection, and accurate ladle identification, ultimately paving the way for practical implementation in industrial environments.

2 Ladle Surface Monitoring

The measurements for working on the thermal characterization of the ladles were made with a Flir a655sc microbolometric camera [7]. Microbolometer FPAs can be created from metal or semiconductor materials and operate according to non-quantum principles. This means that they respond to radiant energy in a way that causes a change of state in the bulk material (i.e., the bolometer effect). Generally, microbolometers do not require cooling, which allows compact camera designs that are relatively low in cost [8]. Thermal images were captured at the billet casting machine turret position. The turret holds ladles for casting and allows quick ladle changes using its two arms. While one ladle is being cast, an empty ladle can be loaded from the opposite side. The thermal camera measures the ladle in this opposite position. During casting, the ladle is filled with hot steel, and after emptying, it appears empty in the camera. The system consists of acquisition, data extraction, and data analysis nodes.

The acquisition node captures data from the thermal camera and filters the sequences to store only necessary and non-redundant information. It monitors a specific directory where the camera saves thermal sequences and related data in a format specific to the camera and software (e.g., ".seq" files generated by ResearchIR software from FLIR). Whenever a new file is added to the directory, the acquisition node algorithm processes it. The algorithm converts the raw data of each sequence into a 3D array, F(i,j,t), with dimensions determined by the camera's pixel resolution and the number of frames captured over time. This stage selects the optimal frame for post-processing based on two criteria. First, the frame must have a maximum temperature exceeding 400°C to ensure a ladle is present in the camera's field of view. Second, the chosen frame minimizes the absolute mean optical flow per frame. This criterion accounts for the stationary period of the ladle after the crane positions it in the tower, as well as its rest period before being lifted again after the casting process.

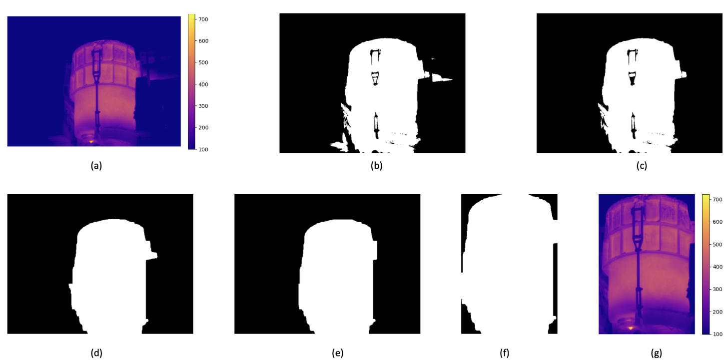

Fig. 1. Ladle recognition algorithm

After the selection of the most appropriate frame, a pre-processing has been developed to eliminate areas that are not of interest for the posterior analysis, such as background, reflections or defective pixels. The steps taken are described below and the effect of each step can be appreciated in Figure 1. The algorithm takes the frame selected on previous paragraph. Every pixel with a temperature value below of 150 C^{o} is discarded. The greatest continuous area is detected, and every other area is discarded. Afterwards, some morphological transformations are applied. A 50 iterations dilation with a 5x5 pixel is followed by an erosion of the same characteristics. Every pixel is discarded if the non-masked pixel amount in its row or column does not surpass the thresholding value of 60 pixels. The mask is cropped so that only the relevant area of the image is saved. The output image is resized with the same dimensions as the mask.

Thanks to this pre-processing, the ladle is centred in the stored image, and it is ready be sent to the data extraction node for obtaining the information within the image and posterior data analysis. Besides this, the total amount of information transmitted between nodes is reduced to the minimum necessary, optimizing times in the processing. With this stage, the acquisition node concludes its work, reading a file in a specific format, and returning the resized image, mask and the metadata stored in the recording file.

The objective of the data extraction node is to extract data from the input provided by the acquisition node for posterior data analysis. The extracted data is stored within the image, but for communication purposes, it is necessary to reduce the data amount to avoid saturating the architecture. The refractory degradation depends on the material it comes in contact with. Three zones can be distinguished in a ladle: the air zone, the slag zone, and the barrel zone. The ladle's nerves are used to differentiate these zones. The algorithm then identifies the hottest spots in each zone. The information from the acquisition node and the ladle's life cycle is combined and stored in a JSON file, significantly reducing the data size. To establish a unified coordinate system, a center is determined for each image, which serves as a common reference point. Artificial vision techniques are employed to locate the nerves or interfaces between the different zones. Finally, the extracted information, including data from the sequence file, caster-provided data, segmentation details, and thermal data, is stored in the JSON file, including temperature values, spot localization, and other relevant information.

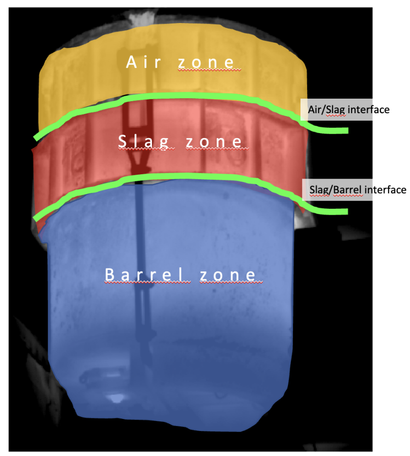

Fig. 2. Graphical description of the ladle's zones

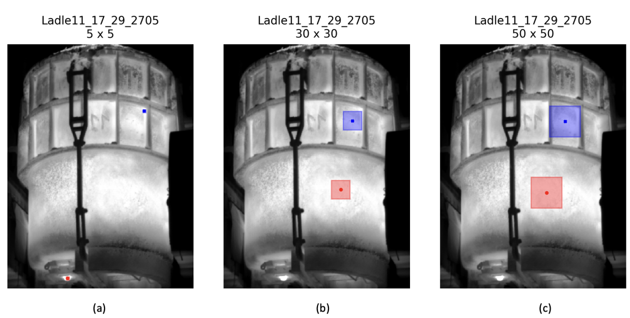

Fig. 3. Different size hottest areas for unique measurement in the slag zone (red) and in the barrel zone (blue). (a) 5x5 area, (b) 30x30 area and (c) 50x50 area

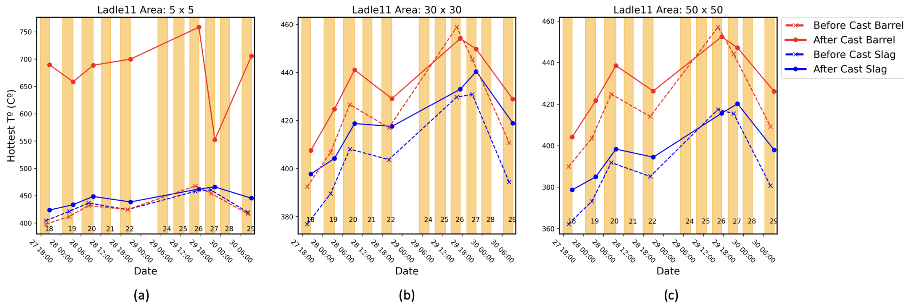

Fig. 4. Ladle 11 temperature analysis for areas (a) 5x5, (b) 30x30 and (c)50x50

Observing the 5x5 chart (Figure [reffig5a]), the effect of molten steel in the bottom of the ladle (barrel zone) is evident, with noticeably higher temperature values compared to other curves. However, there is an anomaly in the curve (use 27) where the temperature measured was lower but still influenced by molten steel. Analyzing the other three curves in Figure 9a, two main conclusions can be drawn. First, the slag zone slightly heats up during the casting process, as indicated by higher after-casting temperatures. Second, temperature values in the slag zone are vaguely higher than in the barrel zone (pre-casting), possibly due to more significant slag erosion compared to molten steel. Figures 4 a-c confirm that temperatures after casting are slightly higher. Vertical profiles can provide more information about the casting process than individual temperature spots.

When considering larger areas (Figure 4 b-c), the previously mentioned patterns are not visible. Barrel temperatures are higher than those in the slag zone, with a greater difference in larger areas. Two reasons explain this: first, the larger areas partially cover the colder nerves, which are colder than the rest of the ladle surface. Second, temperature distribution is more homogeneous in the barrel zone, while the slag zone has hot spots closer to colder areas, leading to counterproductive effects on the overall mean temperature of the hottest area.

3 Ladle Tracking

Although several methods [2] [9], [3] [1] and datasets [4] were introducted for number detection, they are not suitable to be used in steel factories domain because of the image appearance diffeence. In steel factories, the environment is dark, dusty, and usually very hot. Color cameras typically have higher resolution compared to thermal cameras. Initially, the detection algorithm was developed using thermal images, and later sample color images were examined. The color images revealed challenges in detecting the ladle number due to low contrast with the ladle surface, especially in a harsh, dusty environment. Additionally, using RGB color cameras raises concerns about worker security and data privacy in the steel factory setting. Investigating the thermal images revealed the following observations. The resolution of the thermal camera is smaller compared to

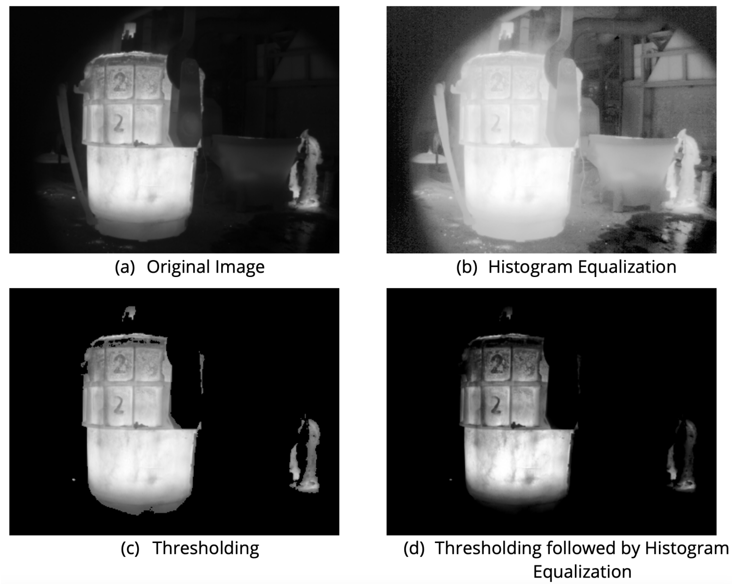

Fig. 5. Thermal Data preprocessing pipeline.

color cameras, with a resolution of 388x288 pixels in the installed camera at the steel factory cleaning station. Despite the smaller size, the numbers can still be detected due to the temperature contrast between the number and the ladle's surface. In contrast, the provided RGB images pose difficulty in identifying the ladles. One thermal image in the figure shows two identifiable digits, with the digit shown in green and the surrounding area in red or a mix of green and red. The algorithm's objective is to locate and identify the number on the ladle's surface.

The ladle is the main object of interest in the camera's view and the hottest spot. The ladle's number is welded on its surface at least once, and the contrast difference between the ladle's surface and the number is important. Image preprocessing techniques were applied to enhance this contrast. Figure 5 shows the results of different pre-processing techniques applied to thermal images. Figure 5 illustrates the image pre-processing techniques used on the thermal data. The original heat map is shown in subfigure (a), which is represented in grayscale as a one-channel image to capture heat variation. Histogram equalization (subfigure (b)) enhances details in the image, revealing background information of the cleaning station that was not visible in the original image. To eliminate this background data, thresholding is applied (subfigure (c)), squashing the corresponding pixels to zero. The final image is obtained by applying histogram equalization to the thresholded image, enhancing the contrast between the ladle's number and the surrounding (colder) pixels. This contrast enhancement is crucial for ladle number identification. At the cleaning station, a thermal camera is installed with an adjusted perspective to acquire a better view of the ladles. The output of the camera is a two-dimensional array containing pixel temperatures. The initial

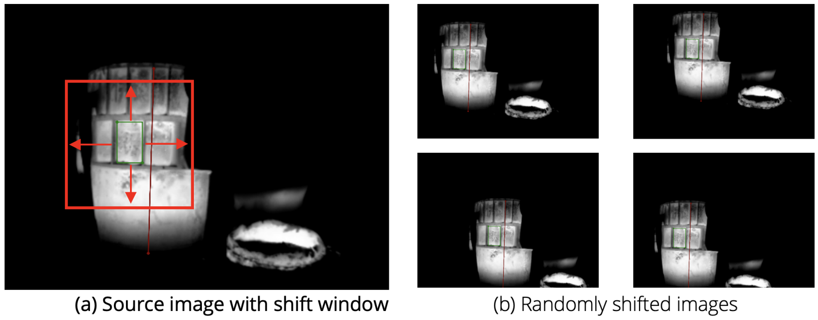

Fig. 6. Synthetic data generation pipeline

data provided for the project was from the old perspective, which was sufficient for initial experiments but not suitable for model validation. The thermal data is a one-channel, two-dimensional image, and generating a heat map results in a three-channel RGB image, regardless of the color scales used.

Thermal data consisting of approximately 20,000 images covering a period of around 3 months was shared. Each image was checked, and those containing a visible number were annotated with the number's bounding box and class (e.g., class 2 for ladle number 2). This annotation process resulted in a dataset of approximately 1,200 images. However, the data had an imbalance issue. Ladle 2 appeared approximately 700 times, while ladle 12 appeared only 49 times. To address this, synthetic data was generated to validate the developed method. The synthesized images were created to ensure a balanced representation of all classes. The specific ladle numbers used in the steel factory (ladles 1, 2, 8, 10, 11, 12, 14, 15, 16, and 17) were taken into account during the data synthesis process.

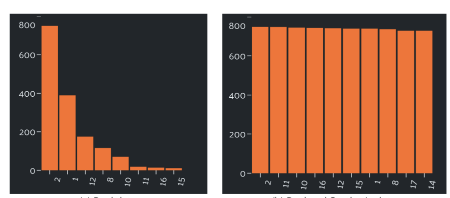

Fig. 7. Ladle images histograms

Twenty seven images were selected to be used as background images. The 27 images were annotated with a location to add the digit at, and a line consisting of two points to represent the orientation of the ladle. The angle of the line can be computed and used to rotate the number before resizing and adding it at the bounding box location. Since the number is colder than the surrounding surface. The intensity of the number was set to the reduced value of the average intensity of the bounding box. Having 27 images would provide a good variation in the background. However, it is not sufficient to generate hundreds of images to try to balance the real data with the synthetic ones. Consequently, image translation was applied on the whole background image.

The ladle tracking task can be decomposed into two main problems, number detection in the thermal image, and number classification of the detected number. In the proposed solution, the numbers are modelled as object, thus making it an object detection task. In the proposed solution, and since there are a limited number of ladles, each ladle number is considered a class. Consequently in the developed model, the data has 10 classes. There are several methods that exist in literature for object detection. One of the state of the are methods is the Faster RCNNs. This methods belongs to the region proposal-based methods for object detection. The method has two deep neural networks, one for each sub task. The first one is for proposing the region where the number could exist in. The second network is for the number classification. The Faster-RCNNs can be implemented using several networks as backbone. In the proposed implementation, Residual Networks ResNet-50 was employed. The training is done in a supervised learning-based fashion. The labelled data is used to train the networks. The annotations either from the real or the synthetic data are in the training.

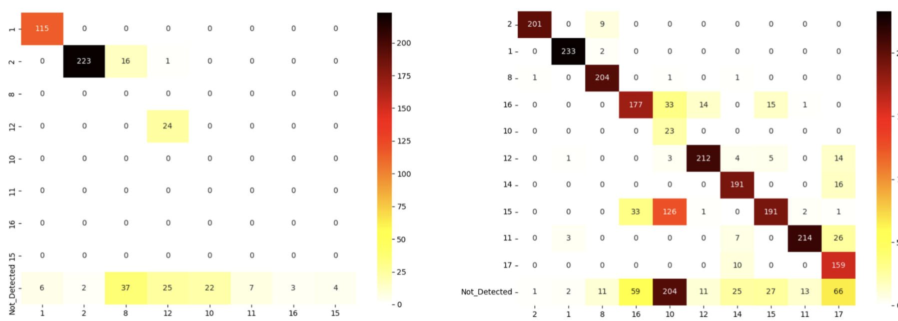

Fig. 8. Confusion Matrix (a) Real Data, (b) Real and Synthetic Data

Experiments and Results Initial experiments were conducted using real data, which consisted of approximately 1200 annotated images. The dataset was divided into 70% for training and 30% for testing. Faster-RCNN Region-Proposal-Network was utilized with pre-trained weights. Training for 1000 epochs yielded satisfactory accuracy, considering the relatively small dataset. The experiment results, depicted in Figure 8, present a confusion matrix plot of the model's detections on the test set. Ladles 1 and 2 were consistently identified correctly, but accuracy decreased for other classes due to insufficient training data. Consequently, as mentioned in the previous subsection, synthetic data was generated and the dataset was augmented. The dataset is balanced. In total, it contains 7100 image. Similar to the previous experiment, the training set used in this experiment was 70% of the data and the test set was 30%. With a significant increase in data volume compared to the previous experiment, the number of iterations was raised to 4000 epochs. The results are presented in Figure 20, showing that almost all classes are accurately detected and identified. The confusion matrix highlights the False Negative detections in the last row. Additionally, an accuracy measurement was computed using a test set consisting of approximately 2200 real and synthetic images. A ladle is considered correctly detected if at least one number on the ladle is accurately detected, even if a second number is present but goes undetected. The proposed method achieved an accuracy of 80%, indicating that around 1760 images were correctly identified by the trained model. However, there appears to be an issue in detecting ladle 10, possibly due to the low resolution of the number in the image. Initial experiments were conducted using approximately 1200 annotated images, with a 70% training set and a 30% testing set. Faster-RCNN Region-Proposal-Network with pre-trained weights was employed and trained for 1000 epochs. Despite the relatively small dataset, satisfactory accuracy was achieved. The experiment results, shown in Figure [refconfusion], display a confusion matrix plot of the model's detections on the test set. Ladles 1 and 2 were consistently identified correctly, but accuracy decreased for other classes due to insufficient training data. To address this, synthetic data was generated and augmented, resulting in a balanced dataset of approximately 7100 images. Sample detections can be seen in Figure 21. The confusion matrix highlights the presence of False Negative detections in the last row. Additionally, an accuracy measurement using a test set of approximately 2200 real and synthetic images yielded an 80% accuracy, indicating that around 1760 images were correctly identified by the trained model. However, there appears to be an issue in detecting ladle 10, possibly due to the low resolution of the number in the image.

4 Conclusion

The work carried out during this study has quantified some information that steelmakers knew beforehand but were proofed thanks to the thermal camera. It was shown that, for the example given with ladle analysed in paragraph 3.3, the slag zone suffers greater erosion due to the chemical corrosion of the slag and therefore offers greater surface temperatures. Besides that, before and after casting differences were analysed and the temperature difference between both images was studied, proofing that the ladle surface remains heating while the casting process, as this process is long and by no means instantaneous.

In the ladle tracking it is shown that it is possible to detect and identify the ladles in the thermal images at the steel factory using computer vision and deep learning-based methods. The use of thermal images was employed, and favored over RGB images in such harsh environment. Data pre-processing was presented for reducing the unnecessary information in the image and enhancing the important section in the thermal image. Synthetic data was presented to balance the dataset. The underlying task was modelled as an object detection problem. The deep learning method used was the state-of-the-art Faster RCNN method.

Acknowledgment

This work has been funded by the EU project HyperCOG (Grant agreement number 869886).

References

1. Boonnag, C., Ittichaiwong, P., Saengmolee, W., Seesawad, N., Chinkamol, A., Rattanasomrerk, S., Veerakanjana, K., Thanontip, K., Limpornchitwilai, W., Wilaiprasitporn, T.: Pacman: a framework for pulse oximeter digit detection and reading in a low-resource setting. IEEE Internet of Things Journal (2023)
2. Liu, H., Bhanu, B.: Pose-guided r-cnn for jersey number recognition in sports. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0 (2019)
3. Maglo, A., Orcesi, A., Pham, Q.C.: Efficient tracking of team sport players with few game-specific annotations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3461–3471 (2022)
4. Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A.Y.: Reading digits in natural images with unsupervised feature learning (2011)
5. Papavasileiou, A., Aivaliotis, P., Aivaliotis, S., Makris, S.: An optical system for identifying and classifying defects of metal parts. International Journal of Computer Integrated Manufacturing **35**(3), 326–340 (2022)
6. Stavropoulos, P., Papacharalampopoulos, A., Sabatakakis, K.: Robust and secure quality monitoring for welding through platform-as-a-service: A resistance and submerged arc welding study. Machines **11**(2), 298 (2023)
7. Systems, F.: Flir a655sc | thermal imaging camera | flir systems (Accessed August, 2023), https://www.flir.eu/products/a655sc
8. Yadav, P.K., Yadav, I., Ajitha, B., Rajasekar, A., Gupta, S., Ashok Kumar Reddy, Y.: Advancements of uncooled infrared microbolometer materials: A review. Sensors and Actuators A: Physical **342** (2022). https://doi.org/https://doi.org/10.1016/j.sna.2022.113611, https://www.sciencedirect.com/science/article/pii/S0924424722002497
9. Zhang, R., Wu, L., Yang, Y., Wu, W., Chen, Y., Xu, M.: Multi-camera multi-player tracking with deep player identification in sports video. Pattern Recognition **102**, 107260 (2020)